

CLAIMS

What is claimed is:

1. A method of caching data for a direct access storage device having a plurality of addressable locations, comprising the steps of:

storing, in a cache memory, copies of equally sized bands of data retrieved from said direct access storage device, each said band comprising a predetermined number of contiguous addressable locations on said direct access storage device spanning a plurality of sectors on said direct access storage device,

responding to a request for access to a storage device location for which a copy is stored in the cache memory, by accessing the copy stored in the cache memory, and

responding to a request for access to a storage device location for which a copy is not stored in the cache memory, by retrieving data for a band of said predetermined number of contiguous addressable locations on said direct access storage device, the band including said desired storage device location and contiguous addressable locations adjacent thereto,

storing said retrieved data for said retrieved band into said cache memory, and

responding to the request for access to the desired storage device location by delivering a portion of the data that was retrieved from the storage device and stored in the cache memory.

2. The method of claim 1 further comprising monitoring accesses to resident bands of data for which

RECEIVED - DECEMBER 1967

copies are stored in the cache memory by maintaining statistics on types of accesses made to bands of data.

3. The method of claim 2 wherein maintaining statistics on types of accesses made to bands of data comprises maintaining a counter associated with each monitored band of data, said counters being credited or penalized in response to types of accesses made to the associated band of data.

4. A method of caching data for a direct access storage device having a plurality of addressable locations, comprising the steps of:

storing, in a cache memory, copies of data retrieved from said direct access storage device,

monitoring accesses to resident data for which copies are stored in the cache memory, and to nonresident data for which copies are not stored in the cache memory,

responding to a request for access to a storage device location for which a copy is stored in the cache memory, by accessing the copy stored in the cache memory, and

responding to a request for access to a storage device location for which a copy is not stored in the cache memory, by determining, based on monitored accesses, whether data for said storage device location for which a copy is not stored in the memory, should be made resident, and if not, accessing the location on said direct access storage device.

retrieving data for a predetermined number of contiguous addressable locations on said direct access storage device, including said desired storage device

location and contiguous addressable locations adjacent thereto,

identifying least advantageous data for which a copy is stored in the cache memory, based on previously monitored accesses to data,

storing said retrieved data for said contiguous addressable locations, in said cache memory in place of the copy in said cache memory of said least advantageous data, and

responding to the request for access to the desired storage device location by delivering a portion of the data that was retrieved from the storage device and stored in the cache memory.

10. The method of claim 4 further comprising maintaining a cache directory identifying data in said direct access storage device for which copies are stored in said cache memory, and identifying memory locations in said cache memory where each said copy is stored.

11. A method of caching data for a direct access storage device having a plurality of addressable locations, comprising the steps of:

storing, in a cache memory, copies of data retrieved from said direct access storage device,

monitoring accesses to resident blocks of data for which copies are stored in the cache memory, and maintaining a resident LRU queue in which resident blocks of data are ordered from most to least recently used, and further maintaining statistics on types of accesses made to resident blocks of data in the form of a counter associated with each monitored block of data,

said counters being credited or penalized in response to types of accesses made to the associated data,

responding to a request for access to a storage device location for which a copy is stored in the cache memory, by accessing the copy stored in the cache memory, and

responding to a request for access to a storage device location for which a copy is not stored in the cache memory, by accessing the location on said direct access storage device.

12. The method of claim 11 wherein maintaining statistics further comprises crediting said counter by a predetermined credit in response to a read to a block of data associated with said counter, and penalizing said counter by a predetermined penalty in response to a write to a block of data associated with said counter.

13. The method of claim 11 wherein accesses are monitored for nonresident blocks of data for which a copy is not stored in the cache memory, by maintaining a nonresident LRU queue in which nonresident blocks of data are ordered from most to least recently used, and further maintaining statistics on types of accesses made to nonresident blocks of data in the form of a counter associated with each monitored block of data, said counters being credited or penalized in response to types of accesses made to the associated data.

14. The method of claim 13 wherein accesses are monitored for at least half as many nonresident blocks as resident blocks.

11-22-69 10:55 AM

15. The method of claim 13 wherein responding to a request for access to a desired storage device location for which a copy is not stored in the cache memory, further comprises

determining whether accesses are being monitored for a nonresident block including said desired storage device location, and if so, based on said statistics, determining whether a copy of said desired storage device location would advantageously be stored in said cache memory, and if so identifying a least advantageous resident block of data, and storing said retrieved data for said desired storage device location, in said cache memory in place of the copy in said cache memory of said least advantageous data.

16. The method of claim 15 wherein determining whether a copy of said desired storage device location would advantageously be stored in said cache memory and identifying least advantageous resident block of data, comprises determining whether statistics for said least recently used resident block of data exceed an aging threshold.

17. The method of claim 16 wherein when statistics for said least recently used resident block of data exceed said aging threshold, said retrieved data for said desired storage location is not stored in said cache memory, said least recently used resident block of data is made most recently used, and its statistics are penalized.

18. The method of claim 16 wherein when statistics for said least recently used resident block of data exceed said aging threshold, it is determined whether statistics of a next least recently used resident block exceed said aging threshold, and if not said next least recently used resident block is identified as the least advantageous resident block of data, and replaced with said retrieved data for said desired storage location.

19. The method of claim 15 further comprising, when accesses are not being monitored for said desired storage device location, initiating statistics for a nonresident block of data including said desired storage device location, and marking said nonresident block including said desired storage device location most recently used. *u*

20. The method of claim 15 wherein determining whether a copy of said desired storage device location would advantageously be stored in said cache memory, comprises comparing statistics for a nonresident block of data including said desired storage device location, to a threshold value.

Sub B2 21. A method of caching data for a direct access storage device having a plurality of addressable locations, comprising the steps of:

storing, in a cache memory, copies of data retrieved from said direct access storage device, and a cache directory identifying data in said direct access storage device for which copies are stored in said cache

memory, and identifying memory locations in said cache memory where each said copy is stored,

responding to a request for access to a storage device location for which a copy is stored in the cache memory, by accessing the copy stored in the cache memory,

responding to a request for access to a desired storage device location for which a copy is not stored in the cache memory, by accessing said desired storage device location from said storage device, and

responding to a change in an amount of cache memory available by altering said cache directory.

22. The method of claim 21 wherein in response to addition of new cache memory, the method further comprises modifying said cache directory to identify memory locations in said new cache memory where copies data from said direct access storage device may be stored, and then storing, in said new cache memory, copies of data retrieved from said direct access storage device.

23. The method of claim 21 wherein in response to the removal of removed cache memory, the method further comprises modifying said cache directory to no longer identify memory locations in said removed cache memory.

24. The method of claim 21 further comprising monitoring accesses to data for which copies are stored in the cache memory.

25. The method of claim 24 wherein monitoring accesses to data comprises maintaining a LRU queue in which data are ordered from most to least recently used.

26. The method of claim 25 wherein monitoring accesses to data comprises maintaining statistics on types of accesses made to data.

27. The method of claim 26 wherein maintaining statistics on types of accesses made to data comprises maintaining a counter associated with blocks of data, said counters being credited or penalized in response to types of accesses made to the associated block of data.

28. The method of claim 27 wherein maintaining statistics further comprises crediting a counter by a predetermined credit in response to a read to a block of data associated with said counter, and penalizing said counter by a predetermined penalty in response to a write to a block of data associated with said counter.

29. The method of claim 27 further comprising identifying a least advantageous block of data for which a copy is stored in the cache memory, based on previously monitored accesses to blocks of data, and as part of responding to a request for access to a desired storage device location for which a copy is not stored in the cache memory, retrieving from said direct access storage device a block of data including said desired storage device location, storing the block of data retrieved from said storage device, in place of

the copy in said cache memory of said least advantageous block of data.

30. The method of claim 24 further comprising monitoring accesses to data for which copies are not stored in the cache memory.

31. A cache control circuit of caching data for a direct access storage device having a plurality of addressable locations, the control circuit performing the steps of:

storing, in a cache memory, copies of data retrieved from said direct access storage device,

monitoring accesses to resident data for which copies are stored in the cache memory, and to nonresident data for which copies are not stored in the cache memory,

responding to a request for access to a storage device location for which a copy is stored in the cache memory, by accessing the copy stored in the cache memory, and

responding to a request for access to a storage device location for which a copy is not stored in the cache memory, by determining, based on monitored accesses, whether data for said storage device location for which a copy is not stored in the memory, should be made resident, and if not, accessing the location on said direct access storage device.

32. The cache control circuit of claim 31 wherein monitoring accesses to data comprises

maintaining a LRU queue in which data are ordered from most to least recently used.

33. The cache control circuit of claim 31 wherein monitoring accesses to data comprises maintaining statistics on types of accesses made to data.

34. The cache control circuit of claim 33 wherein said data is arranged into block and accesses to data are monitored on a block-by-block basis, and maintaining statistics on types of accesses made to data comprises maintaining a counter associated with each monitored block of data, said counters being credited or penalized in response to types of accesses made to the associated block of data. *u*

35. The cache control circuit of claim 34 wherein maintaining statistics further comprises crediting said counter by a predetermined credit in response to a read to a block of data associated with said counter, and penalizing said counter by a predetermined penalty in response to a write to a block of data associated with said counter.

35. A cache control circuit of caching data for a direct access storage device having a plurality of addressable locations, the control circuit performing the steps of:

storing, in a cache memory, copies of data retrieved from said direct access storage device,

monitoring accesses to resident blocks of data for which copies are stored in the cache memory, and maintaining a resident LRU queue in which resident blocks of data are ordered from most to least recently used, and further maintaining statistics on types of accesses made to resident blocks of data in the form of a counter associated with each monitored block of data, said counters being credited or penalized in response to types of accesses made to the associated data,

responding to a request for access to a storage device location for which a copy is stored in the cache memory, by accessing the copy stored in the cache memory, and

responding to a request for access to a storage device location for which a copy is not stored in the cache memory, by accessing the location on said direct access storage device. *a*

36. The cache control circuit of claim 35 wherein maintaining statistics further comprises crediting said counter by a predetermined credit in response to a read to a block of data associated with said counter, and penalizing said counter by a predetermined penalty in response to a write to a block of data associated with said counter.

37. The cache control circuit of claim 35 wherein accesses are monitored for nonresident blocks of data for which a copy is not stored in the cache memory, by maintaining a nonresident LRU queue in which nonresident blocks of data are ordered from most to least recently used, and further maintaining statistics

on types of accesses made to nonresident blocks of data in the form of a counter associated with each monitored block of data, said counters being credited or penalized in response to types of accesses made to the associated data.

38. The cache control circuit of claim 37 wherein accesses are monitored for at least half as many nonresident blocks as resident blocks.

39. A program product, comprising:

(a) a program configured to perform a method of caching data for a direct access storage device having a plurality of addressable locations, the method comprising:

storing, in a cache memory, copies of data retrieved from said direct access storage device,
monitoring accesses to resident data for which copies are stored in the cache memory, and to nonresident data for which copies are not stored in the cache memory,
responding to a request for access to a storage device location for which a copy is stored in the cache memory, by accessing the copy stored in the cache memory, and
responding to a request for access to a storage device location for which a copy is not stored in the cache memory, by determining, based on monitored accesses, whether data for said storage device location for which a copy is not stored in the memory,

100-100000-100000

(b) a signal bearing media bearing the program.

41. The program product of claim 39, wherein the signal bearing media is a recordable media.

(a) a program configured to perform a method of caching data for a direct access storage device having a plurality of addressable locations, the method comprising:

Page 67 of 69
IBM R0998-0164
WH&E IBM/67
Patent Application

responding to a request for access to a storage device location for which a copy is stored in the cache memory, by accessing the copy stored in the cache memory, and responding to a request for access to a storage device location for which a copy is not stored in the cache memory, by accessing the location on said direct access storage device; and

(b) a signal bearing media bearing the program.

43. The program product of claim 42, wherein the signal bearing media is a transmission type media.

44. The program product of claim 42, wherein the signal bearing media is a recordable media.

ADD B3 >

TOP SECRET